

PARAMETRIC INFERENCE FOR PROPORTIONAL (REVERSE) HAZARD RATE MODELS WITH NOMINATION SAMPLING

Mohammad Nourmohammadi, Mohammad Jafari Jozani¹, and Brad C. Johnson

University of Manitoba, Department of Statistics, Winnipeg, MB, CANADA, R3T 2N2

Abstract

Randomized nomination sampling (RNS) is a rank-based sampling technique which has been shown to be effective in several nonparametric studies involving environmental and ecological applications. In this paper, we investigate parametric inference using RNS design for estimating the unknown vector of parameters θ in the proportional hazard rate and proportional reverse hazard rate models. We examine both maximum likelihood (ML) and method of moments (MM) methods and investigate the relative precision of our proposed RNS-based estimators compared with those based on simple random sampling (SRS). We introduce four types of RNS-based data as well as necessary EM algorithms for the ML estimation, and evaluate the performance of corresponding estimators in estimating θ . We show that there are always values of the design parameters on which RNS-based estimators are more efficient than those based on SRS. Inference based on imperfect ranking is also explored and it is shown that the improvement holds even when the ranking is imperfect. Theoretical results are augmented with numerical evaluations and a case study.

Keywords: Randomized nomination sampling; Method of moments; Maximum likelihood; Modified maximum likelihood; Proportional hazard rate; Proportional reverse hazard rate; EM algorithm.

1 Introduction

Randomized nomination sampling (RNS) is a rank-based sampling scheme. Rank-based sampling schemes are data collection techniques which utilize the advantage of additional information available in the population to provide an artificially stratified sample with more structure. Providing more structured sample enables us to direct our attention toward units that represent the underlying population. Let X be an absolutely continuous random variable distributed according to the cumulative distribution function (CDF) $F(x; \theta)$ and the probability density function (PDF) $f(x; \theta)$, where F is known and $\theta = (\theta_1, \dots, \theta_p)^T \in \Theta \subseteq \mathbb{R}^p$ (p -dimensional Euclidean space), is unknown. Further, let $\{K_i : i \in \mathbb{N}\}$ be a sequence of independent random variables taking values in \mathbb{N} (the natural numbers) with probabilities $\rho = \{(\rho_1, \rho_2, \dots) : \sum_{i=1}^{\infty} \rho_i = 1\}$ so that $\mathbb{P}(K_i = j) = \rho_j$, $j \in \{1, 2, \dots\}$. Let $\{Z_i : i \in \mathbb{N}\}$ be a sequence of independent Bernoulli random variables with success probability $\zeta \in [0, 1]$, independent of K_i and X . The RNS design consists of drawing m random sample sets of size K_i , $i = 1, \dots, m$, from the population for the purpose of ranking and finally nominating m sampling units (one from each set) for

¹Corresponding author. Email: m_jafari_jozani@umanitoba.ca. Tel: 1 204 272 1563.

final measurement. The nominee from the i -th set is the largest unit of the set with probability ζ or the smallest unit of the set with probability $1 - \zeta$. The RNS observation Y_i can be written as

$$Y_i = Z_i X_{K_i:K_i} + (1 - Z_i) X_{1:K_i}, \quad i = 1, \dots, m,$$

where $X_{1:K_i}$ and $X_{K_i:K_i}$ are respectively the smallest and the largest units of the i^{th} set of size K_i .

RNS was introduced by Jafari Jozani and Johnson (2012) and applied to the problem of estimating population total in finite population. Later, this sampling scheme was applied in constructing confidence intervals for quantiles in finite populations (Nourmohammadi et al., 2014) and infinite populations (Nourmohammadi et al., 2015b), as well as in constructing tolerance intervals (Nourmohammadi et al., 2015a). Some well-known examples of RNS are given below:

- (1) The choice of $K_i = 1$, $i = 1, \dots, m$, results in the SRS design with observations denoted by X_i , $i = 1, \dots, m$.
- (2) The choice of $\zeta = 1$ nominates the maximum from each set and results in a maxima nomination sampling (MANS) design (see Willemain, 1980; Boyles and Samaniego, 1986; Tiwari, 1988; Kvam and Samaniego, 1993; Tiwari and Wells, 1989; Jafari Jozani and Mirkamali, 2010; and Jafari Jozani and Mirkamali, 2011).
- (3) The choice of $\zeta = 0$ nominates the minimum from each set and results in a minima nomination sampling (MINS) design (see Tiwari and Wells, 1989 and Wells et al., 1990).
- (4) The choice of $\zeta = \frac{1}{2}$ and $k_i = k$, for a constant $k \in \mathbb{N}$, results in a randomized extreme ranked set sampling (RERSS) design (see Jafari Jozani and Johnson, 2012).
- (5) The choice of $Z_{2i-1} = 1$ and $Z_{2i} = 0$, and $K_i = k$, where $i = 1, \dots, m$, for a constant $k \in \mathbb{N}$ and an even number m , results in an extreme ranked set sampling (ERSS) design (see Samawi et al., 1996, and Ghosh and Tiwari, 2009).
- (6) The choice of $Z_{2i-1} = 1$ and $Z_{2i} = 0$, and $K_i = i$, where $i = 1, \dots, m$, for an even number m results in a moving extreme ranked set sampling (MRSS) design (see Al-Odat and Al-Saleh, 2001).

Note that Z_i , $i = 1, \dots, m$, in (5) and (6) are no longer independent and identically distributed (IID). RNS is a cost-effective method of selecting data in situations in which measuring the characteristic of interest is difficult, expensive and/or destructive, but a small number of sampling units can be cheaply ranked and the minimum and the maximum observations can be easily identified. Unlike the regular ranked set sampling (RSS), RNS allows for an increase of the set size without introducing too much ranking error. Identifying only the extremes, rather than providing a complete ranking on the units in the set, is more practical, since we need to identify successfully only the first or the last ordered unit. Regarding the randomness of the set size, while the RNS technique allows one to select the sets of fixed size k , i.e. $\mathbb{P}(K_i = k) = 1$ for some fixed k and $i = 1, \dots, m$, providing the flexibility of choosing the

sizes in random helps to apply the RNS scheme in circumstances where the set size might be random (see Boyles and Samaniego (1986) and Gemayel et al. (2010)). Another advantage of allowing the set size to be random is that, when $\rho_1 > 0$, randomized nomination sample is expected to contain a simple random sample of size $m\rho_1$ in addition to a collection of extremal order statistics from various set sizes, which contain more information about the underlying population than SRS. RNS also has the flexibility to adjust the proportion of maximums and minimums by choosing an appropriate value for $\zeta \in [0, 1]$ based on the population shape. This reduces the concern in working with extremes when the underlying population is skewed.

The RNS-based statistical inference may be made under various situations. For example, there might be the case where y_i 's are the only available information and no further information is provided on either k_i or z_i , $i = 1, \dots, m$. There might also be situations in which the size of sets or the number of maximums (and subsequently the number of minimums), or both are chosen in advance, instead of getting involved in a randomized process. In the situation where k_i and/or z_i are known, the CDF of Y_i can be found by conditioning on k_i and z_i , or both. The conditioning argument makes the theoretical investigation more complicated, but it provides more efficient statistical inference. In this paper, both unconditional and conditional RNS are studied. Four types of RNS data are introduced, corresponding to situations where, for any set $i = 1, \dots, m$, (1) the triplet (y_i, k_i, z_i) are all known, (2) only (y_i, k_i) are known, (3) only (y_i, z_i) are known, or (4) only y_i are known. These types of RNS data are, respectively, called RNS complete-data, Type I, Type II, and Type III RNS incomplete-data.

We discuss RNS-based maximum likelihood (ML) and method of moments (MM) estimates of the population parameters when the underlying random variable follows the proportional hazard rate (PHR) or proportional reverse hazard rate (PRHR) model. Let F_0 be an absolutely continuous probability distribution function with density f_0 , possibly depending on an unknown vector of parameters $\boldsymbol{\theta}$, and let $c = \sup\{x \in \mathbb{R} : F_0(x) = 0\}$ and $d = \inf\{x \in \mathbb{R} : F_0(x) = 1\}$ (where, by convention, we take $\inf \emptyset = -\sup \emptyset = \infty$) so that, if $X \sim F_0$, we have $-\infty \leq c \leq X \leq d \leq \infty$ with $F_0(c) = 1 - F_0(d) = 0$. The family of PHR models (based on F_0) is given by the family of distributions

$$F(x; \boldsymbol{\theta}) = 1 - [1 - F_0(x)]^{1/\gamma(\boldsymbol{\theta})}, \quad (1)$$

where $\gamma(\boldsymbol{\theta}) > 0$. Similarly, the family of PRHR models is given by the family of distributions

$$F(x; \boldsymbol{\theta}) = [F_0(x)]^{1/\gamma(\boldsymbol{\theta})}. \quad (2)$$

The hazard rate function $H(x)$ and the reverse hazard rate function $R(x)$ at $x \in (c, d)$ are given respectively by

$$H(x) = \frac{1}{\gamma(\boldsymbol{\theta})} \frac{f_0(x)}{1 - F_0(x)} \quad \text{and} \quad R(x) = \frac{1}{\gamma(\boldsymbol{\theta})} \frac{f_0(x)}{F_0(x)}. \quad (3)$$

The PHR and PRHR models in (1) and (2) are well-known in lifetime experiments. The lifetime distribution of a system and its components are of interest in reliability testing. Statistical analysis

Table 1: Some examples of hazard and reverse hazard rate models

Distribution	$F(x; \boldsymbol{\theta})$	Domains	$F_0(x)$	$\gamma(\boldsymbol{\theta})$	Rate Function
$X \sim \text{Exp}(\lambda)$	$1 - e^{-x/\lambda}$	$x \in [0, \infty), \lambda \in (0, \infty)$	$1 - e^{-x}$	λ	$1/\lambda$
$X \sim \text{Par}(\nu = 1, \lambda)$	$1 - x^{-1/\lambda}$	$x \in [1, \infty), \lambda \in (0, \infty)$	$1 - x^{-1}$	λ	$-1/\lambda x$
$X \sim \text{Bet}(1/\eta, \nu = 1)$	$x^{1/\eta}$	$x \in [0, 1], \eta \in (0, \infty)$	x	$1/\eta$	$1/\eta x$
$X \sim \text{GExp}(1/\eta, \nu = 1)$	$(1 - e^{-x})^{1/\eta}$	$x \in [0, \infty), \eta \in (0, \infty)$	$1 - e^{-x}$	η	$e^{-x}/\eta(1 - e^{-x})$

of the lifetime of a system or its components is an important topic in many research areas such as engineering, marketing and biomedical sciences. See, for example, Lawless (2011), Navarro et al. (2008), Gupta and Gupta (2007), Gupta et al. (1998) and Helsen and Schmittlein (1993). The PHR and PRHR models include several well-known lifetime distributions. In the sequel, we are interested in estimating $\gamma(\boldsymbol{\theta})$ with specific choices of F_0 . Some examples of hazard and reverse hazard rate models are presented in Table 1.

The remainder of the paper is organized as follows. In Section 2, we investigate the RNS complete-data and provide the PDF and CDF of an RNS observation in the form of complete-data. We also derive the ML estimators of $\gamma(\boldsymbol{\theta})$ in the PHR and PRHR model when the triplet (y_i, k_i, z_i) , $i = 1, \dots, m$, is available. In Section 3, we present the ML estimation for the parameters based on incomplete RNS data. We provide the PDF and CDF of observations in each RNS incomplete-data and use the EM algorithm to obtain the ML estimators of the parameters of interest. In Section 4, we derive the RNS-based MM estimation in the PHR and PRHR models; when the RNS data are from either complete- or incomplete-data scenarios. In Section 5, we illustrate the numerical results in detail and compare the performance of the RNS-based estimators with the corresponding SRS estimators for the Exponential and Beta distributions. Moreover, in Section 5, the performance of RNS-based ML estimators in a more complicated situation is investigated using a real life dataset on fish mercury contamination measurement.

2 ML Estimation in RNS Complete-Data

Let X_1, \dots, X_m be a simple random sample of size m from a continuous distribution with CDF $F(x; \boldsymbol{\theta})$ and PDF $f(x; \boldsymbol{\theta})$. If it exists, the SRS-based ML estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_s$, satisfies the ML equations

$$\sum_{i=1}^m \frac{f'(X_i; \boldsymbol{\theta})}{f(X_i; \boldsymbol{\theta})} = 0, \quad (4)$$

where $f'(X; \boldsymbol{\theta}) = \partial f(X; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. Let Y_1, \dots, Y_m be a randomized nomination sample of size m from F . The forms of the CDFs and PDFs of Y_i 's, in addition to the RNS-based ML equations, are determined by the available RNS data. In this section, we use the RNS complete-data to derive the ML estimator of $\boldsymbol{\theta}$. In the RNS complete-data case, the triplets (y_i, k_i, z_i) , $i = 1, 2, \dots, m$, are known. In other words, one knows that, for $i = 1, \dots, m$, the observed value $Y_i = y_i$ is from a set of size $K_i = k_i$ with the value

$Z_i = z_i$ and the rank $r_i = z_i k_i + (1 - z_i)$ in the i -th set, where k_i and z_i are both known. An RNS observation Y given $K = k$ and $Z = z$, where $k \in \{1, 2, \dots\}$ and $z \in \{0, 1\}$, has the CDF $G_c(y|k, z; \boldsymbol{\theta})$ and the PDF $g_c(y|k, z; \boldsymbol{\theta})$ as follows

$$G_c(y|k, z; \boldsymbol{\theta}) = \{F^k(y; \boldsymbol{\theta})\}^z \{1 - \bar{F}^k(y; \boldsymbol{\theta})\}^{1-z},$$

and

$$g_c(y|k, z; \boldsymbol{\theta}) = k f(y; \boldsymbol{\theta}) \{F^{k-1}(y; \boldsymbol{\theta})\}^z \{\bar{F}^{k-1}(y; \boldsymbol{\theta})\}^{1-z}.$$

The log likelihood function based on the RNS complete-data is given by

$$L_c^{RNS}(\boldsymbol{\theta}) = \sum_{i=1}^m \{\log k_i + \log f(y_i; \boldsymbol{\theta}) + z_i(k_i - 1) \log F(y_i; \boldsymbol{\theta}) + (1 - z_i)(k_i - 1) \log \bar{F}(y_i; \boldsymbol{\theta})\}. \quad (5)$$

Upon differentiation of (5) with respect to $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ and equating the result to zero, the (complete) ML estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_c$, is obtained from

$$\frac{\partial L_c^{RNS}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m \left\{ \frac{f'(y_i; \boldsymbol{\theta})}{f(y_i; \boldsymbol{\theta})} + z_i(k_i - 1) \frac{F'(y_i; \boldsymbol{\theta})}{F(y_i; \boldsymbol{\theta})} + (1 - z_i)(k_i - 1) \frac{\bar{F}'(y_i; \boldsymbol{\theta})}{\bar{F}(y_i; \boldsymbol{\theta})} \right\} = 0, \quad (6)$$

where $F'(y; \boldsymbol{\theta}) = \partial F(y; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. Since both $F'(y; \boldsymbol{\theta}) / F(y; \boldsymbol{\theta})$ and $\bar{F}'(y; \boldsymbol{\theta}) / \bar{F}(y; \boldsymbol{\theta})$ are involved in the RNS likelihood, equation (6) is more complicated to solve for $\boldsymbol{\theta}$ than (4), and for most distributions there is no closed form expressions for the ML estimators. Following the idea proposed by Mehrotra and Nanda (1974), we consider the modified ML (MML) estimators of parameters. Depending on the underlying distribution, one may need to replace one or both of the second and third terms on the left-hand side of (6) by their corresponding expected values. The obtained MML estimator of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}_m$.

2.1 ML Estimation in the PHR Model

Let X_i , $i = 1, \dots, m$, be a sequence of IID random variables from the family of PHR models in (1). The SRS-based ML estimator of $\gamma(\boldsymbol{\theta})$, denoted by $\widehat{\gamma_s(\boldsymbol{\theta})}$, can be expressed as

$$\widehat{\gamma_s(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \log \bar{F}_0(X_i), \quad (7)$$

which is an unbiased estimator of $\gamma(\boldsymbol{\theta})$ with variance $\text{Var}[\widehat{\gamma_s(\boldsymbol{\theta})}] = \gamma^2(\boldsymbol{\theta})/m$. Under the model (1), the PDF of a random variable Y_i , $i = 1, \dots, m$, from the RNS complete-data is

$$g_c(y_i|k_i, z_i; \boldsymbol{\theta}) = k_i \frac{1}{\gamma(\boldsymbol{\theta})} \frac{f_0(y_i)}{\bar{F}_0(y_i)} \left([\bar{F}_0(y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}} \right)^{\alpha_i - 1} \left(1 - [\bar{F}_0(y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}} \right)^{\beta_i}, \quad (8)$$

where $\alpha_i = (1 - z_i)(k_i - 1) + 2$ and $\beta_i = z_i(k_i - 1)$. The RNS-based complete ML estimator of $\gamma(\boldsymbol{\theta})$, denoted by $\hat{\gamma}_c(\boldsymbol{\theta})$, is obtained by solving the equation

$$\sum_{i=1}^m \left\{ \gamma(\boldsymbol{\theta}) + \log \bar{F}_0(Y_i) - \beta_i \frac{[\bar{F}_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}}{1 - [\bar{F}_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}} \log \bar{F}_0(Y_i) + (\alpha_i - 2) \log \bar{F}_0(Y_i) \right\} = 0. \quad (9)$$

Note that if $z_i = 0$, for all $i = 1, \dots, m$, one can easily obtain a complete RNS-based ML estimator of $\gamma(\boldsymbol{\theta})$ as described in the following lemma.

Lemma 2.1. *Let X_1, \dots, X_m be IID random variables from (1), and suppose $(Y_1, k_1, z_1=0), \dots, (Y_m, k_m, z_m=0)$ is the corresponding MINS sample of size m . The complete ML estimator of $\gamma(\boldsymbol{\theta})$ is given by*

$$\widehat{\gamma_c(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m k_i \log \bar{F}_0(Y_i), \quad (10)$$

which is an unbiased estimator of $\gamma(\boldsymbol{\theta})$ with the variance equal to its SRS-based counterpart, i.e.,

$$\text{Var}[\widehat{\gamma_c(\boldsymbol{\theta})}] = \frac{\gamma^2(\boldsymbol{\theta})}{m}.$$

Proof. From (9), by replacing $z_i = 0$ for $i = 1, \dots, m$, the complete ML estimate for $\gamma(\boldsymbol{\theta})$ in (10) is obtained. Noting that in the PHR model we have

$$\text{E}[\log \bar{F}_0(Y_i)] = k_i B(\alpha_i - 1, \beta_i + 1) \text{E}(\log W_i) \gamma(\boldsymbol{\theta}), \quad (11)$$

and

$$\text{E}[\log^2 \bar{F}_0(Y_i)] = k_i B(\alpha_i - 1, \beta_i + 1) \text{E}(\log^2 W_i) \gamma^2(\boldsymbol{\theta}), \quad (12)$$

where $W_i \sim \text{Beta}(\alpha_i - 1, \beta_i + 1)$, $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ and

$$\text{E}[\log W_i] = \mathfrak{D}(\alpha_i - 1) - \mathfrak{D}(\alpha_i + \beta_i),$$

and

$$\text{Var}[\log W_i] = \mathfrak{D}'(\alpha_i - 1) - \mathfrak{D}'(\alpha_i + \beta_i).$$

The function $\mathfrak{D}(\cdot)$ is the Digamma function and, for $n \in \mathbb{N}$, $\mathfrak{D}(n) = \sum_{j=1}^{n-1} \frac{1}{j} - \tau$, where $\tau \approx 0.57722$ is the Euler-Mascheroni constant. The function $\mathfrak{D}'(\cdot)$ is the Trigamma function, which is defined as

$$\mathfrak{D}'(n) = -\int_0^1 \frac{x^{n-1} \log x}{1-x} dx, \quad \text{for } n \in \mathbb{N},$$

where $\mathfrak{D}'(n+1) = \mathfrak{D}'(n) - 1/n^2$ gives the result. Now, the expected value and the variance of $\widehat{\gamma_c(\boldsymbol{\theta})}$ follow immediately. \square

In the general case, to construct the MML estimator of $\gamma(\boldsymbol{\theta})$ in PHR models, one needs to replace the second term in (9) by its expected value for any $z_i \neq 0$ and $k_i \neq 1$. Note that for $z_i = 0$ or $k_i = 1$ the second term in (9) equals zero. The RNS-based MML estimator of $\gamma(\boldsymbol{\theta})$, denoted by $\widehat{\gamma_m(\boldsymbol{\theta})}$, and the corresponding expected value and variance are given in the following theorem.

Theorem 2.2. Let X_1, \dots, X_m be IID random variables from (1), and suppose $(Y_1, k_1, z_1), \dots, (Y_m, k_m, z_m)$ is the corresponding RNS sample of size m with at least one $z_i = 1$. Further, let $U_i \sim \text{Beta}(\alpha_i, \beta_i)$ and $W_i \sim \text{Beta}(\alpha_i - 1, \beta_i + 1)$, where $\alpha_i = (1 - z_i)(k_i - 1) + 2$, $\beta_i = z_i(k_i - 1)$ and

$$\mathcal{E}_i = \frac{-1}{\gamma(\boldsymbol{\theta})} \mathbb{E} \left(\frac{[\bar{F}_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}}{1 - [\bar{F}_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}} \log \bar{F}_0(Y_i) \right) = -k_i B(\alpha_i, \beta_i) \mathbb{E}(\log U_i), \quad i = 1, \dots, m.$$

(a) The MML estimator of $\gamma(\boldsymbol{\theta})$ based on RNS complete-data of size m is given by

$$\widehat{\gamma_m(\boldsymbol{\theta})} = \frac{-\sum_{i=1}^m (\alpha_i - 1) \log \bar{F}_0(Y_i)}{m + \sum_{i=1}^m \beta_i \mathcal{E}_i}, \quad (13)$$

(b) The expected value and the variance of $\hat{\gamma}_m(\boldsymbol{\theta})$ are respectively

$$\mathbb{E}[\widehat{\gamma_m(\boldsymbol{\theta})}] = \frac{-\sum_{i=1}^m (\alpha_i - 1) \mathbb{E}[\log \bar{F}_0(Y_i)]}{m + \sum_{i=1}^m \beta_i \mathcal{E}_i} \gamma(\boldsymbol{\theta}), \quad (14)$$

and

$$\text{Var}[\widehat{\gamma_m(\boldsymbol{\theta})}] = \frac{\sum_{i=1}^m (\alpha_i - 1)^2 \text{Var}(\log \bar{F}_0(Y_i))}{(m + \sum_{i=1}^m \beta_i \mathcal{E}_i)^2} \gamma^2(\boldsymbol{\theta}), \quad (15)$$

where $\mathbb{E}[\log \bar{F}_0(Y_i)]$ and $\text{Var}[\log \bar{F}_0(Y_i)]$ are obtained from (11) and (12).

Proof. For part (a), the value \mathcal{E}_i is derived using the PDF of Y_i in (8). Substituting \mathcal{E}_i for the second term in (9) results in (13). Parts (b) is trivial. \square

Remark 2.3. Considering (14),

$$\widehat{\gamma_m(\boldsymbol{\theta})} = \left[\frac{m + \sum_{i=1}^m \beta_i \mathcal{E}_i}{-\sum_{i=1}^m (\alpha_i - 1) k_i B(\alpha_i - 1, \beta_i + 1) \mathbb{E}(\log W_i)} \right] \widehat{\gamma_m(\boldsymbol{\theta})}$$

is an unbiased estimator of $\gamma(\boldsymbol{\theta})$.

Corollary 2.4. The MML estimator of $\gamma(\boldsymbol{\theta})$ based on a MANS sample of size m from (1) is given by

$$\widehat{\gamma_m(\boldsymbol{\theta})} = -\frac{1}{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j}} \sum_{i=1}^m \log \bar{F}_0(Y_i),$$

which is an unbiased estimator of $\gamma(\boldsymbol{\theta})$ with variance

$$\text{Var}[\widehat{\gamma_m(\boldsymbol{\theta})}] = \frac{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j^2}}{(\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j})^2} \gamma^2(\boldsymbol{\theta}), \quad (16)$$

which is always smaller than its SRS counterpart.

Example 2.5. Let $X_i, i = 1, \dots, m$, be a SRS sample of size m from an exponential distribution with parameter λ . Further, let (Y_i, k_i, z_i) be an RNS sample of size m obtained from the same exponential distribution. Noting that

$$F(x; \lambda) = 1 - e^{-x/\lambda} \text{ and } f(x; \lambda) = \frac{1}{\lambda} e^{-x/\lambda},$$

the SRS-based ML estimator of λ is $\hat{\lambda}_s = \bar{X}$ with the expected value $E[\hat{\lambda}_s] = \lambda$ and the variance $\text{Var}[\hat{\lambda}_s] = \lambda^2/m$. Assuming $W_i \sim \text{Beta}(\alpha_i - 1, \beta_i + 1)$, where $\alpha_i = (1 - z_i)(k_i - 1) + 2$ and $\beta_i = z_i(k_i - 1)$, we have $E[Y_i] = -\lambda k_i B(\alpha_i - 1, \beta_i + 1) E(\log W_i)$ and $E(Y_i^2) = \lambda^2 k_i B(\alpha_i - 1, \beta_i + 1) E(\log^2 W_i)$. The RNS-based complete ML estimator of λ , denoted by $\hat{\lambda}_c$, is obtained from the equation

$$\sum_{i=1}^m \left\{ \lambda - (\alpha_i - 1)Y_i + \beta_i \frac{e^{-Y_i/\lambda}}{1 - e^{-Y_i/\lambda}} Y_i \right\} = 0.$$

The MINS complete-data ML estimator of λ is given by

$$\hat{\lambda}_c = \frac{\sum_{i=1}^m k_i Y_i}{m},$$

which is unbiased for λ with the variance $\text{Var}[\hat{\lambda}_c] = \lambda^2/m$. Also, the MML estimator of λ based on MANS complete-data is

$$\hat{\lambda}_m = \frac{\sum_{i=1}^m Y_i}{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j}}. \quad (17)$$

Noting that $E[Y_i] = \lambda \sum_{j=1}^{k_i} \frac{1}{j}$ and $\text{Var}[Y_i] = \lambda^2 \sum_{j=1}^{k_i} \frac{1}{j^2}$, the MML estimator $\hat{\lambda}_m$ in (17) is unbiased with variance

$$\text{Var}[\hat{\lambda}_m] = \frac{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j^2}}{(\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j})^2} \lambda^2.$$

Remark 2.6. It can be shown that the above mentioned results hold with minor modifications for the family of PRHR model in (2). For example, the RNS-based complete ML estimator of the parameter $\gamma(\boldsymbol{\theta})$, is obtained from

$$\sum_{i=1}^m \left\{ \gamma(\boldsymbol{\theta}) + \log F_0(Y_i) + \beta_i \log F_0(Y_i) - (\alpha_i - 2) \left(\frac{[F_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}}{1 - [F_0(Y_i)]^{\frac{1}{\gamma(\boldsymbol{\theta})}}} (\log F_0(Y_i)) \right) \right\} = 0. \quad (18)$$

Also, the RNS-based complete ML estimator of $\gamma(\boldsymbol{\theta})$ in (2) using MANS sample is given by

$$\widehat{\gamma_c(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m k_i \log F_0(Y_i),$$

which is an unbiased estimator of $\gamma(\boldsymbol{\theta})$ with variance equal to its SRS-based counterpart, i.e., λ^2/m . Note that the ML estimator of $\gamma(\boldsymbol{\theta})$ in PRHR models based on SRS is given by

$$\widehat{\gamma_s(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \log F_0(X_i).$$

The MML estimator of $\gamma(\boldsymbol{\theta})$ in the case where at least one $z_i = 0$ is given by

$$\widehat{\gamma_m(\boldsymbol{\theta})} = \frac{-\sum_{i=1}^m (\beta_i + 1) \log F_0(Y_i)}{m + \sum_{i=1}^m (\alpha_i - 2) \mathcal{E}_i^*},$$

where $\mathcal{E}_i^* = -k_i B(\beta_i + 2, \alpha_i - 2) E(\log U_i^*)$, and $U_i^* \sim \text{Beta}(\beta_i + 2, \alpha_i - 2)$. The expected value and the variance of $\widehat{\gamma_m(\boldsymbol{\theta})}$ are, respectively,

$$E[\widehat{\gamma(\boldsymbol{\theta})}] = \frac{-\sum_{i=1}^m (\beta_i + 1) k_i B(\beta_i + 1, \alpha_i - 1) E[\log E_i^*]}{m + \sum_{i=1}^m (\alpha_i - 2) \mathcal{E}_i^*} \gamma(\boldsymbol{\theta}) \quad (19)$$

and

$$\text{Var}[\widehat{\gamma(\boldsymbol{\theta})}] = \frac{\sum_{i=1}^m (\beta_i + 1)^2 \text{Var}[\log \bar{F}_0(Y_i)]}{(m + \sum_{i=1}^m (\alpha_i - 2) \mathcal{E}_i^*)^2} \gamma^2(\boldsymbol{\theta}), \quad (20)$$

where, assuming $W_i^* \sim \text{Beta}(\beta_i + 1, \alpha_i - 1)$, we have

$$E[\log F_0(Y_i)] = \gamma(\boldsymbol{\theta}) k_i B(\beta_i + 1, \alpha_i - 1) E(\log W_i^*), \quad (21)$$

and

$$E[\log^2 F_0(Y_i)] = \gamma^2(\boldsymbol{\theta}) k_i B(\beta_i + 1, \alpha_i - 1) E(\log^2 W_i^*). \quad (22)$$

In the MINS design, the MML estimator of $\gamma(\boldsymbol{\theta})$ is given by

$$\widehat{\gamma_m(\boldsymbol{\theta})} = \frac{-\sum_{i=1}^m \log F_0(Y_i)}{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j}}.$$

One can easily show that $\widehat{\gamma_m(\boldsymbol{\theta})}$ is an unbiased estimator of $\gamma(\boldsymbol{\theta})$ with the variance given by (16).

Example 2.7. Let X_1, \dots, X_m be an SRS sample of size m from a $\text{Beta}(\frac{1}{\eta}, 1)$ distribution with the corresponding CDF and PDF as follow

$$F(x; \boldsymbol{\theta}) = x^{\frac{1}{\eta}} \quad \text{and} \quad f(x; \eta) = \frac{1}{\eta} x^{\frac{1}{\eta}-1}. \quad (23)$$

Further, let $Y_i, i = 1, \dots, m$, denote an RNS sample from (23). The RNS-based complete MLE, $\hat{\eta}_c$, is obtained from the equation

$$\sum_{i=1}^m \left\{ \eta + \log y_i + z_i(k_i - 1) \log Y_i + (1 - z_i)(k_i - 1) \frac{Y_i^{\frac{1}{\eta}}}{1 - Y_i^{\frac{1}{\eta}}} (-\log Y_i) \right\} = 0.$$

For a MANS sample, the unbiased RNS-based ML estimate for η using the complete-data is given by

$$\hat{\eta}_c = -\frac{\sum_{i=1}^m k_i \log Y_i}{m},$$

with $\text{Var}(\hat{\eta}_c) = \eta^2/m$. Also, based on the MINS design, the parameters α_i and β_i are, respectively, 1 and k_i . It is seen that the unbiased MML estimator of η based on RNS complete-data is given by

$$\hat{\eta}_m = \frac{-\sum_{i=1}^m \log Y_i}{\sum_{i=1}^m \sum_{j=1}^{k_i} \frac{1}{j}},$$

with the variance provided in (16).

3 ML Estimation in RNS Incomplete-Data

In RNS complete-data the triplet (y_i, k_i, z_i) are all assumed to be observed. In practice, for $i = 1, \dots, m$, one of k_i or z_i , or both, may be unknown. In this section, we investigate the CDF and PDF of the RNS random variable Y_i , the ML equations and corresponding EM algorithms associated with each type of RNS incomplete-data.

(a) **Type I RNS incomplete-data:** Here, we assume that (y_i, k_i) , for $i = 1, \dots, m$, are known. In other words, the i -th observed unit is from a set of size k_i . This unit may be the maximum or minimum with probability ζ or $1 - \zeta$, respectively. The CDF $G_1(y|k; \boldsymbol{\theta})$ and PDF $g_1(y|k; \boldsymbol{\theta})$ for the observed values $Y = y$ given $K = k$ are, respectively, as follow

$$\begin{aligned} G_1(y|k; \boldsymbol{\theta}) &= \zeta F^k(y; \boldsymbol{\theta}) + (1 - \zeta)(1 - \bar{F}^k(y; \boldsymbol{\theta})), \text{ and} \\ g_1(y|k; \boldsymbol{\theta}) &= k f(y; \boldsymbol{\theta}) \{ \zeta F^{k-1}(y; \boldsymbol{\theta}) + (1 - \zeta) \bar{F}^{k-1}(y; \boldsymbol{\theta}) \}. \end{aligned}$$

The log likelihood function based on $\mathbf{y} = (y_1, y_2, \dots, y_m)$ and $\mathbf{k} = (k_1, k_2, \dots, k_m)$ is given by

$$L_1^{RNS} = \sum_{i=1}^m \log k_i + \sum_{i=1}^m \log f(y_i; \boldsymbol{\theta}) + \sum_{i=1}^m \log \{ \zeta F^{k_i-1}(y_i; \boldsymbol{\theta}) + (1 - \zeta) \bar{F}^{k_i-1}(y_i; \boldsymbol{\theta}) \}. \quad (24)$$

Upon differentiating (24) with respect to $\boldsymbol{\theta}$ and equating the result to zero, we have

$$\begin{aligned} \frac{\partial L_1^{RNS}}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^m \frac{f'(y_i; \boldsymbol{\theta})}{f(y_i; \boldsymbol{\theta})} \\ &+ \sum_{i=1}^m \left\{ \frac{\zeta (k_i - 1) F^{k_i-2}(Y_i; \boldsymbol{\theta}) F'(Y_i; \boldsymbol{\theta}) + (1 - \zeta) (k_i - 1) \bar{F}^{k_i-2}(Y_i; \boldsymbol{\theta}) \bar{F}'(Y_i; \boldsymbol{\theta})}{\zeta F^{k_i-1}(y_i; \boldsymbol{\theta}) + (1 - \zeta) \bar{F}^{k_i-1}(y_i; \boldsymbol{\theta})} \right\} \\ &= 0, \end{aligned}$$

which does not yield explicit solutions for $\hat{\boldsymbol{\theta}}$. In order to find the ML estimators, we use the EM algorithm. In order to pose this problem in the incomplete-data Type I context, we introduce the unobservable data $\mathbf{z} = (z_1, z_2, \dots, z_m)$, where $z_i = 1$ or $z_i = 0$ according to whether the selected unit in the i -th set is the maximum or the minimum, respectively. We find the ML estimator of $\boldsymbol{\theta}$ by adding the unobservable data to the problem via working with the current conditional expectation of the complete-data log likelihood (5) given the observed data and proceed as follow. Let $\boldsymbol{\theta}^{(t)}$ be the value specified for $\boldsymbol{\theta}$ in the t -th iteration. Then on the $(t + 1)$ -th iteration, the conditional expectation of L_c^{RNS} given \mathbf{y} and \mathbf{k} using $\boldsymbol{\theta}^{(t)}$, i.e., $E_{\boldsymbol{\theta}^{(t)}} [L_c^{RNS} | \mathbf{y}, \mathbf{k}]$ is computed. This step is called the E-step. As L_c^{RNS} is a linear function of the unobservable data \mathbf{z} , the E-step is performed simply by replacing z_i by their current conditional expectations given the observed data \mathbf{y} and \mathbf{k} . Therefore, for a known parameter ζ , we have

$$E_{\boldsymbol{\theta}^{(t)}} [Z_i | \mathbf{y}, \mathbf{k}] = \mathbb{P}_{\boldsymbol{\theta}^{(t)}} (Z_i = 1 | y_i, k_i) = z_i^{(t)},$$

where

$$z_i^{(t)} = \frac{g_c(y_i|k_i, z_i = 1, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)})}{g_1(y_i|k_i, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)})} = \frac{\zeta F^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})}{\zeta F^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)}) + (1 - \zeta) \bar{F}^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})}.$$

The next step on the $(t + 1)$ -th iteration, which is called the M-step, requires replacing z_i 's by $z_i^{(t)}$'s in (5) to obtain $\boldsymbol{\theta}^{(t+1)}$ by maximizing $L_c^{RNS}(\boldsymbol{\theta})$. We keep alternating between $z_i^{(t)}$ and $\boldsymbol{\theta}^{(t)}$ until $\boldsymbol{\theta}^{(t)}$ converges to a fixed point.

When the parameter ζ is unknown, the procedure may be started with the initial value of $\zeta^{(0)} \approx 1$ (in PHR model) or $\zeta^{(0)} \approx 0$ (in PRHR model), and continued by updating ζ using $\zeta^{(t+1)} = \frac{1}{m} \sum_{i=1}^m z_i^{(t)}$.

- (b) **Type II RNS incomplete-data:** Here, we consider the case where (y_i, z_i) are known, but the set size k_i is unknown. In other words, we observed the value y_i and we know if the observed unit is the maximum or the minimum unit of the set, but the set size is unknown. The CDF $G_2(y|z; \boldsymbol{\theta})$ and PDF $g_2(y|z; \boldsymbol{\theta})$ for the observed value $Y = y$ given $Z = z$ are, respectively, as follow

$$G_2(y|z; \boldsymbol{\theta}) = \sum_{k=1}^{\infty} \mathbb{P}(K = k) G_c(y|k, z; \boldsymbol{\theta}) = \sum_{k=1}^{\infty} \rho_k \{F^k(y; \boldsymbol{\theta})\}^z \{1 - \bar{F}^k(y; \boldsymbol{\theta})\}^{1-z}, \quad (25)$$

and

$$g_2(y|z; \boldsymbol{\theta}) = f(y, \boldsymbol{\theta}) \sum_{k=1}^{\infty} k \rho_k \{F^{k-1}(y; \boldsymbol{\theta})\}^z \{\bar{F}^{k-1}(y; \boldsymbol{\theta})\}^{1-z}.$$

From (25), the log likelihood function for $\boldsymbol{\theta}$ obtained from the observed data $\mathbf{y} = (y_1, y_2, \dots, y_m)$ and $\mathbf{z} = (z_1, z_2, \dots, z_m)$ is expressed as

$$L_2^{RNS} = \sum_{i=1}^m \log f(y_i, \boldsymbol{\theta}) + \sum_{i=1}^m \log \left[\sum_{k_i=1}^{\infty} k_i \rho_{k_i} \{F^{k_i-1}(y_i, \boldsymbol{\theta})\}^{z_i} \{\bar{F}^{k_i-1}(y_i, \boldsymbol{\theta})\}^{1-z_i} \right]. \quad (26)$$

Upon equating the derivatives of (26) with respect to $\boldsymbol{\theta}$ to zero, we have

$$\begin{aligned} \frac{\partial L_2^{RNS}}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^m \frac{f'(y_i, \boldsymbol{\theta})}{f(y_i, \boldsymbol{\theta})} \\ &+ \sum_{i=1}^m (-1)^{1-z_i} f(y_i, \boldsymbol{\theta}) \left\{ \frac{\sum_{k_i=1}^{\infty} k_i (k_i - 1) \rho_{k_i} \{F^{k_i-2}(y_i, \boldsymbol{\theta})\}^{z_i} \{\bar{F}^{k_i-2}(y_i, \boldsymbol{\theta})\}^{1-z_i}}{\sum_{k_i=1}^{\infty} k_i \rho_{k_i} \{F^{k_i-1}(y_i, \boldsymbol{\theta})\}^{z_i} \{\bar{F}^{k_i-1}(y_i, \boldsymbol{\theta})\}^{1-z_i}} \right\} \\ &= 0, \end{aligned}$$

which apparently do not yield explicit solutions for the incomplete-data MLE of $\boldsymbol{\theta}$. Since the vector $\mathbf{k} = (k_1, k_2, \dots, k_m)$ is unobservable, we are unable to estimate $\boldsymbol{\theta}$ by the maximizing (5). So we again use the EM algorithm. In the E-step, we substitute the unobservable data in (5) by averaging the complete-data log likelihood over its conditional distribution given the observed \mathbf{y} and \mathbf{z} . As L_c^{RNS} is a linear function of the unobservable data \mathbf{k} , the E-step is performed simply by replacing

k_i by their current conditional expectations given the observed data \mathbf{y} and \mathbf{z} . Therefore, for a known parameter $\boldsymbol{\rho}$,

$$\mathbb{E}_{\boldsymbol{\theta}^{(t)}}[K_i|\mathbf{y}, \mathbf{z}] = \sum_{k=1}^{\infty} k \mathbb{P}_{\boldsymbol{\theta}^{(t)}}(K_i = k|\mathbf{y}, \mathbf{z}) = k_i^{(t)},$$

where

$$\begin{aligned} k_i^{(t)} &= \sum_{k=1}^{\infty} k \rho_k \frac{g_c(y_{i,t}|z_i, k_i = k, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)})}{g_2(y_{i,t}|z_i, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)})} \\ &= \frac{\sum_{k=1}^{\infty} k^2 \rho_k \{F^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})\}^{z_i} \{\bar{F}^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})\}^{1-z_i}}{\sum_{k=1}^{\infty} k \rho_k \{F^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})\}^{z_i} \{\bar{F}^{k_i-1}(y_i; \boldsymbol{\theta}^{(t)})\}^{1-z_i}}. \end{aligned}$$

The M-step on the $(t+1)$ -th iteration requires replacing k_i 's by $k_i^{(t)}$'s in (5) to obtain $\boldsymbol{\theta}^{(t+1)}$ by maximizing $L_c^{RNS}(\boldsymbol{\theta})$. We keep alternating between $k_i^{(t)}$ and $\boldsymbol{\theta}^{(t)}$ until $\boldsymbol{\theta}^{(t)}$ converges to a fixed point.

When the parameter $\boldsymbol{\rho}$ is unknown, the procedure can be started with the initial value $\rho_1^{(0)} = (1/n, \dots, 1/n)$, where the length of $\rho_1^{(0)}$ is a relatively large but arbitrary n , and continued by calculating $\rho_{k_i}^{(t+1)}$ using the frequencies of the $k_i^{(t)}$ over m .

- (c) **Type III RNS incomplete-data:** Here, we study the case where only y_i is observed and no more information on the set size and the rank of the selected unit is available. The CDF $G_3(y|\boldsymbol{\theta})$ and PDF $g_3(y|\boldsymbol{\theta})$ for the observed value $Y = y$ are given, respectively, by

$$G_3(y; \boldsymbol{\theta}) = \zeta \sum_{k=1}^{\infty} \rho_k F^k(y; \boldsymbol{\theta}) + (1 - \zeta) \sum_{k=1}^{\infty} \rho_k (1 - \bar{F}^k(y; \boldsymbol{\theta})),$$

and

$$g_3(y; \boldsymbol{\theta}) = f(y; \boldsymbol{\theta}) \left\{ \zeta \sum_{k=1}^{\infty} k \rho_k F^{k-1}(y; \boldsymbol{\theta}) + (1 - \zeta) \sum_{k=1}^{\infty} k \rho_k \bar{F}^{k-1}(y; \boldsymbol{\theta}) \right\}.$$

The log likelihood function for $\boldsymbol{\theta}$ formed on the basis of \mathbf{y} is given by

$$L_3^{RNS} = \sum_{i=1}^m \log f(y_i, \boldsymbol{\theta}) + \sum_{i=1}^m \log \left\{ \zeta \sum_{k=1}^{\infty} k \rho_k F^{k-1}(y; \boldsymbol{\theta}) + (1 - \zeta) \sum_{k=1}^{\infty} k \rho_k \bar{F}^{k-1}(y; \boldsymbol{\theta}) \right\}. \quad (27)$$

Upon equating the derivatives of (27) with respect to $\boldsymbol{\theta}$ to zero, the following results are obtained:

$$\begin{aligned} \frac{\partial L_3^{RNS}}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^m \frac{f'(y_i; \boldsymbol{\theta})}{f(y_i; \boldsymbol{\theta})} \\ &+ \sum_{i=1}^m f(y_i; \boldsymbol{\theta}) \left\{ \frac{\zeta \sum_{k=1}^{\infty} k(k-1) \rho_k F^{k-2}(y_i; \boldsymbol{\theta}) - (1 - \zeta) \sum_{k=1}^{\infty} k(k-1) \rho_k \bar{F}^{k-2}(y_i; \boldsymbol{\theta})}{\zeta \sum_{k=1}^{\infty} k \rho_k F^{k-1}(y_i; \boldsymbol{\theta}) + (1 - \zeta) \sum_{k=1}^{\infty} k \rho_k \bar{F}^{k-1}(y_i; \boldsymbol{\theta})} \right\} \\ &= 0. \end{aligned} \quad (28)$$

Similar to Type I and Type II incomplete-data, no explicit ML estimator for the parameter $\boldsymbol{\theta}$ can be obtained from (28). In this type of RNS incomplete-data, two unobservable data sets in the

form of $\mathbf{z} = (z_1, z_2, \dots, z_m)$ and $\mathbf{k} = (k_1, k_2, \dots, k_m)$ are introduced. In order to perform the EM algorithm assuming $\boldsymbol{\rho}$ and ζ are known, we first calculate

$$E_{\boldsymbol{\theta}^{(t)}}[K_i|\mathbf{y}] = \sum_{k=1}^{\infty} k \mathbb{P}_{\boldsymbol{\theta}^{(t)}}(K_i = k|\mathbf{y}) = k_i^{(t)},$$

where

$$k_i^{(t)} = \frac{\sum_{k=1}^{\infty} k^2 \rho_k \{ \zeta F^{k-1}(y_i; \boldsymbol{\theta}^{(t)}) + (1 - \zeta) \bar{F}^{k-1}(y_i; \boldsymbol{\theta}^{(t)}) \}}{\zeta \sum_{k=1}^{\infty} k \rho_k F^{k-1}(y_i; \boldsymbol{\theta}^{(t)}) + (1 - \zeta) \sum_{k=1}^{\infty} k \rho_k \bar{F}^{k-1}(y_i; \boldsymbol{\theta}^{(t)})}.$$

Then, we obtain

$$E_{\boldsymbol{\theta}^{(t)}}[Z_i|\mathbf{y}] = \mathbb{P}_{\boldsymbol{\theta}^{(t)}}(Z_i = 1|\mathbf{y}) = z_i^{(t)},$$

where

$$z_i^{(t)} = \frac{\zeta F^{k_i^{(t)}-1}(y_i; \boldsymbol{\theta}^{(t)})}{\zeta F^{k_i^{(t)}-1}(y_i; \boldsymbol{\theta}^{(t)}) + (1 - \zeta) \bar{F}^{k_i^{(t)}-1}(y_i; \boldsymbol{\theta}^{(t)})}.$$

The M-step on the $(t+1)$ -th iteration requires replacing z_i 's by $z_i^{(t)}$'s and k_i 's by $k_i^{(t)}$'s in (5) to obtain $\boldsymbol{\theta}^{(t+1)}$ by maximizing $L_c^{RNS}(\boldsymbol{\theta})$. We keep alternating between $z_i^{(t)}$, $k_i^{(t)}$ and $\boldsymbol{\theta}^{(t)}$ until $\boldsymbol{\theta}^{(t)}$ converges to a fixed point. When the parameters ζ and $\boldsymbol{\rho}$ are unknown, similar procedures proposed in Type I incomplete-data (for ζ) and in Type II incomplete-data (for $\boldsymbol{\rho}$) are used.

4 RNS-Based MM Estimators

Finding the ML estimators of $\boldsymbol{\theta}$ for complete-data case requires finding the roots of the nonlinear equations (9) and (18), which are cumbersome and computationally expensive. When the available data is incomplete, the iterative EM algorithm for calculating the ML estimator of $\boldsymbol{\theta}$ is not easy-to-use. In this section, we briefly study the MM estimation based on RNS data for $\gamma(\boldsymbol{\theta})$ in PHR and PRHR models. The SRS-based MM estimate of $\gamma(\boldsymbol{\theta})$, denoted by $\widehat{\gamma_{sm}(\boldsymbol{\theta})}$ is equal to the SRS-based ML estimate, $\widehat{\gamma_s(\boldsymbol{\theta})}$, obtained from (7). In PHR, considering the random variable $\log \bar{H}(X)$, the MM estimator of $\gamma(\boldsymbol{\theta})$ can be obtained by equating the first moment of the population to the sample moment as follow

$$\widehat{\gamma_{sm}(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \log \bar{F}_0(X_i).$$

Similarly, in PRHR model, the MM estimator of $\gamma(\boldsymbol{\theta})$ is expressed as

$$\widehat{\gamma_{sm}(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \log F_0(X_i).$$

Now, we present the RNS-based complete- and incomplete-data MM estimators of $\gamma(\boldsymbol{\theta})$ in both PHR and PRHR models.

Table 2: The value of R_i introduced in Theorem 4.1 for RNS complete-data and Type I, Type II and Type II incomplete-data

RNS Data	R_i	
	PHR	PRHR
Complete	$-k_i B(\alpha_i - 1, \beta_i + 1) E[\log W_i]$	$-k_i B(\beta_i + 1, \alpha_i - 1) E[\log W_i^*]$
Type I	$\zeta \sum_{j=1}^{k_i} \frac{1}{j} + (1 - \zeta) \frac{1}{k_i}$	$(1 - \zeta) (\sum_{j=1}^{k_i} \frac{1}{j}) + \zeta \frac{1}{k_i}$
Type II	$-\sum_{k_i=1}^{\infty} \rho_{k_i} k_i B(\alpha_i - 1, \beta_i + 1) E[\log W_i]$	$-\sum_{k_i=1}^{\infty} \rho_{k_i} k_i B(\beta_i + 1, \alpha_i - 1) E[\log W_i^*]$
Type III	$\sum_{k_i=1}^{\infty} \rho_{k_i} \left\{ \zeta \sum_{j=1}^{k_i} \frac{1}{j} + (1 - \zeta) \frac{1}{k_i} \right\}$	$\sum_{k_i=1}^{\infty} \rho_{k_i} \left\{ (1 - \zeta) \sum_{j=1}^{k_i} \frac{1}{j} + \zeta \frac{1}{k_i} \right\}$

Theorem 4.1. *Let Y_1, \dots, Y_m be an RNS sample of size m obtained from a continuous CDF of the family of PHR model or PRHR model. Further, let $W_i \sim \text{Beta}(\alpha_i - 1, \beta_i + 1)$ and $W_i^* \sim \text{Beta}(\beta_i + 1, \alpha_i - 1)$, where $\alpha_i = (1 - z_i)(k_i - 1) + 2$ and $\beta_i = z_i(k_i - 1)$. Then, the unbiased MM estimators of $\gamma(\boldsymbol{\theta})$ in PHR and PRHR models are, respectively, obtained as*

$$\widehat{\gamma_m(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \frac{1}{R_i} \log \bar{F}_0(Y_i) \quad \text{and} \quad \widehat{\gamma_m(\boldsymbol{\theta})} = -\frac{1}{m} \sum_{i=1}^m \frac{1}{R_i} \log F_0(Y_i),$$

where the value of R_i depends on the RNS data type and the underlying model as presented in Table 2.

Note that for the RNS complete-data, the variance of $\widehat{\gamma_m(\boldsymbol{\theta})}$ in both PHR and PRHR models provided in Theorem 4.1 are derived using (11), (12), (21), and (22).

5 Numerical Studies

In this section, we perform numerical studies to compare the performance of the proposed RNS methods with SRS in estimating some parameters. First, we perform some simulations to examine the performance of RNS compared with SRS under different scenarios of available information about the observations, set sizes, and rank of observations. Then, in a case study we evaluate the precision of the RNS design over the SRS design in a more complicated scenario in both perfect and imperfect settings.

5.1 Simulations

We first discuss the reduction in the mean square error (MSE) of the ML estimators in the RNS complete-data in the PHR and PRHR models using the relative precision. The relative precision is defined as the ratio of the RNS-based MSE over the SRS-based MSE such that values less than one are desired. For the incomplete-data settings, the performance of MLE's of the population parameters in two distributions are examined; the parameter λ in the exponential distribution introduced in Example 2.5 and the parameter η in the beta distribution introduced in Example 2.7. Note that the expected value and the variance of the RNS complete-data in the PHR and PRHR models presented in (14), (15),

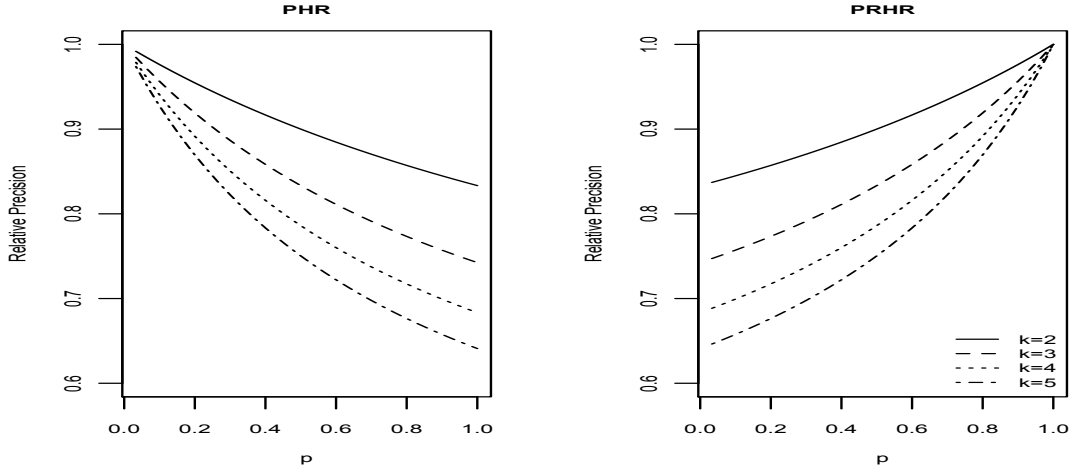


Figure 1: The relative precision of the ML estimators of $\gamma(\theta)$ based on the RNS complete-data over their SRS counterparts in the PHR (left panel) and PRHR (right panel) models when $K = 2, 3, 4, 5$ and the proportion of maximums is $p = 0(0.1)1$. Values less than one show RNS performs better than SRS.

(19), and (20) do not depend on the observed data and are only functions of K and Z . In addition, we investigate the role of the RNS design parameters in improving the performance of the RNS-based estimators compared with their SRS counterparts.

In Figure 1, we provide the MSE of $\widehat{\gamma_c(\theta)}$, the estimator of $\gamma(\theta)$ in the RNS setting when the data is complete, over the MSE of $\hat{\gamma}_s(\theta)$, the estimator of $\gamma(\theta)$ in the SRS setting, for the PHR (left panel) and PRHR (right panel) models. The relative precision is calculated for four RNS designs with fixed set sizes as $K = 2, 3, 4, 5$ and the proportion of maximums varies in $p = 0(0.1)1$. From the results we can compare RNS complete-data with different fixed set sizes and proportion of maximums among themselves and with SRS in terms of the performance of estimators of $\gamma(\theta)$. For example, in the left panel of Figure 1, which shows the relative precision for the PHR models, it is seen that any RNS design, even with $K = 2$ and proportion of maximums $p = 0.1$, outperforms SRS. Increasing the set size and the proportion of maximums improves the performance of the RNS complete-data. The best performance pertains to MANS with $K = 5$. In the right panel of Figure 1, which shows the relative precision for the PRHR models, similar results are obtained except that the best performance pertains to MINS with $K = 5$.

In Figure 2, we provide the relative precision of the ML estimators of λ as the parameter of the exponential distribution in Example 2.5 in three RNS incomplete-data Type I, Type II, and Type III. The relative precision is calculated by the mean square of the RNS-based ML estimate of $\gamma(\theta)$ over its SRS-based counterpart, so values less than one are desired. The top left panel shows the relative precision of the RNS-based ML estimator of λ in the incomplete-data Type I. The relative precision is calculated for the ML estimators of $\lambda = 1, 2, 3, 4$ and $\zeta \in [0, 1]$. It is seen that for the larger values of λ , the RNS incomplete-data Type I outperforms SRS for any $\zeta \in [0, 1]$. As ζ approaches to 1, regardless of the value of the parameter of interest, the performance of RNS incomplete-data Type I becomes better than SRS. The top right panel presents the relative precision of the RNS incomplete-data Type II for

the range of $\lambda = 1(0.1)5$ and for four distributions of the set size K as follows

$$\boldsymbol{\rho}_1 = (0.4, 0.3, 0.2, 0.1), \boldsymbol{\rho}_2 = (0.1, 0.2, 0.3, 0.4), \boldsymbol{\rho}_3 = (0.2, 0, 0, 0.8), \boldsymbol{\rho}_4 = (0, 0, 0, 1).$$

It is seen that the RNS incomplete-data Type II with the assumed $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$, $\boldsymbol{\rho}_3$, and $\boldsymbol{\rho}_4$ improves the precision of the estimators of λ especially when the set sizes get larger. As the value of λ increases, the performance of the RNS incomplete-data Type II is improved more and the distributions of K perform similarly. The next four panels in Figure 2 present the relative precision of the RNS incomplete-data Type III for $\lambda = 1(0.1)5$ and $\zeta \in \{0, 0.25, 0.75, 1\}$. The relative precision for small λ depends on ζ . The last four panels in Figure 2 show that, for all the considered distributions on K , by increasing ζ , RNS outperforms SRS and the relative precision reaches the lowest value when $\zeta = 1$.

In Figure 3 we present the relative precision of the ML estimators of η as the parameter of the beta distribution in the form of Example 2.7 for the RNS incomplete-data Types I, II and III. The top left panel shows the relative precision of the RNS-based ML estimator of $\eta = 1, 2, 3$ and 4 in the incomplete-data Type I for $\zeta \in [0, 1]$. It is seen that for the examined values of η , $\zeta = 0$ improves the RNS incomplete-data Type I over SRS. The top right panel presents the relative precision of the RNS incomplete-data Type II for the range of $\eta = 1(0.1)5$ and for four distributions on K , which are shown by $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$, $\boldsymbol{\rho}_3$, and $\boldsymbol{\rho}_4$. For the examined $\boldsymbol{\rho}$'s, RNS incomplete-data Type II outperforms SRS. The next four panels in Figure 3 present the relative precision of the RNS incomplete-data Type III for $\eta = 1(0.1)5$ and $\zeta \in \{0, 0.25, 0.75, 1\}$. It is seen that for $\zeta = 0$ the RNS incomplete-data Type III performs better than SRS. The relative precision of the estimators obtained from the RNS design with ζ other than zero might works good for some values of η , especially when ζ is close to zero.

We also evaluated the performance of the RNS-based MM estimators of λ and η . Figure 4 shows the precision of $\widehat{\gamma_{cm}(\boldsymbol{\theta})}$, the MM estimators of $\gamma(\boldsymbol{\theta})$ in the RNS setting when the data is complete, relative to their SRS counterparts for the PHR (left panel) and PRHR (right panel) models. The relative precision is calculated for four RNS designs with fixed set sizes when the sets are of sizes $K = 2, 3, 4, 5$ and the proportion of maximums varies in $p = 0(0.1)1$. The results show that the RNS design outperforms SRS for all considered distributions of K ($\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$, $\boldsymbol{\rho}_3$, and $\boldsymbol{\rho}_4$) and for all proportions of maximums $p \in [0, 1]$. We observe in the left panel that, similar to the ML estimators in the RNS-based complete-data, increasing the set size and the proportion of maximums improve the performance of the RNS complete-data in the PHR model. In the PHR model, the best performance is obtained from the MANS design, where all the selected units are maximums, and with the set size $K = 5$. In the PRHR model (right panel), the best performance belongs to the MINS design, where all the selected units are minimums, with the set size $K = 5$.

Figure 5 provides the relative precision of the MM estimators of parameter λ in the exponential distribution introduced in Example 2.5 based on the RNS incomplete-data Type I, Type II, and Type III. The top left panel shows the relative precision of the RNS-based MM estimators in the incomplete-data Type I. It shows that $\zeta = 1$, regardless of the parameter value λ , is the optimum value of ζ which improves the RNS-based MM estimator in the incomplete-data Type I scenario compared with SRS.

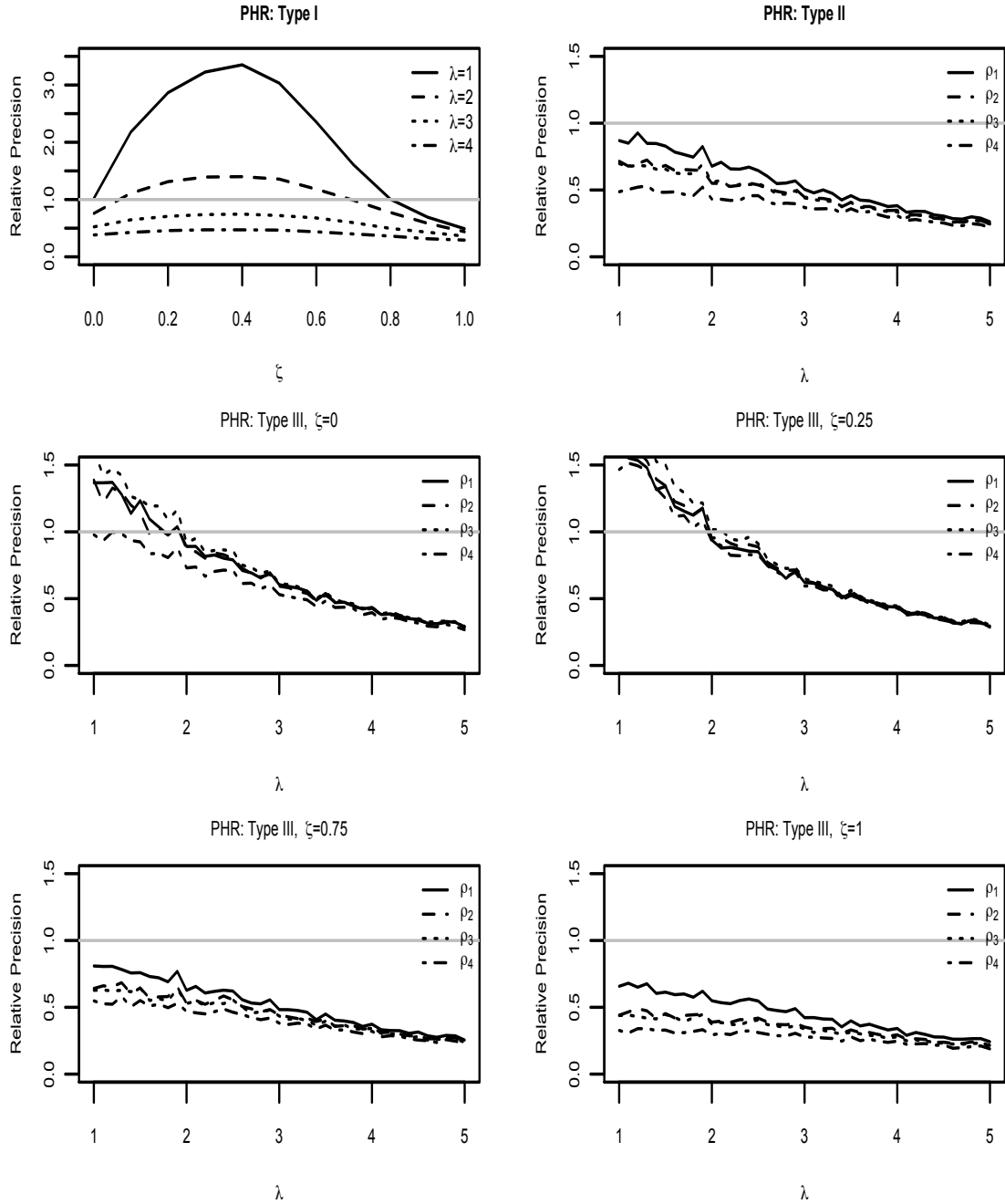


Figure 2: The relative precision of the RNS incomplete-data Type I (top left panel) for $\zeta \in [0, 1]$ and $\lambda \in \{1, 2, 3, 4\}$, Type II (top right panel) for four distributions on K and $\lambda = 1(0.1)5$, and Type III (middle and lower panels) for $\zeta \in \{0, 0.25, 0.75, 1\}$ and $\lambda = 1(0.1)5$ in an exponential distribution with parameter λ and $m = 10$. Values of the relative precision less than one show RNS performs better than SRS.

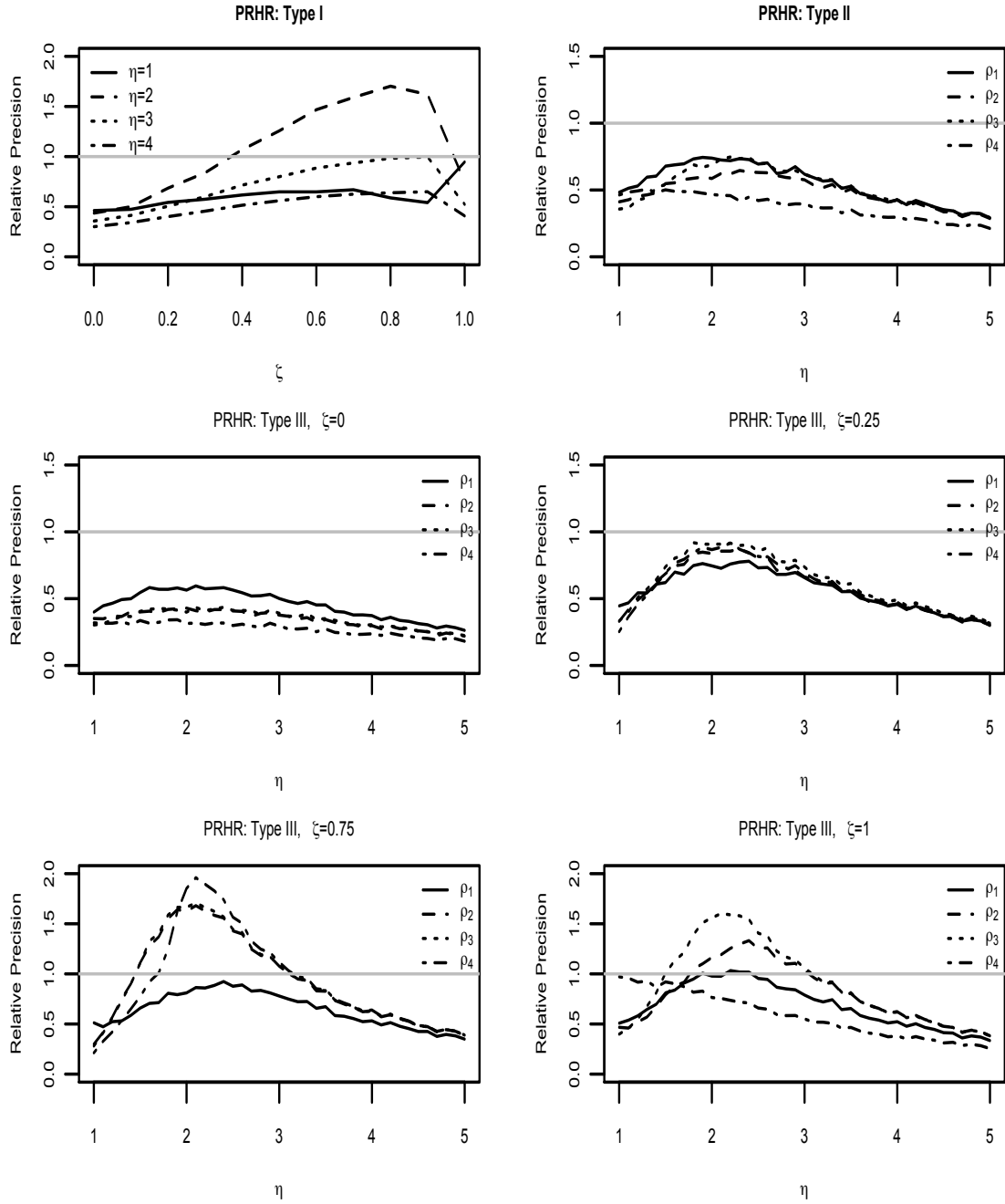


Figure 3: The relative precision of the RNS incomplete-data Type I (top left panel) for $\zeta \in [0, 1]$ and $\eta \in \{1, 2, 3, 4\}$, Type II (top right panel) for four distributions on K and $\eta = 1(0.1)5$, and Type III (middle and lower panels) for $\zeta \in \{0, 0.25, 0.75, 1\}$ and $\eta = 1(0.1)5$ in an exponential distribution with parameter η and $m = 10$. Values of the relative precision less than one shows RNS performs better than SRS.

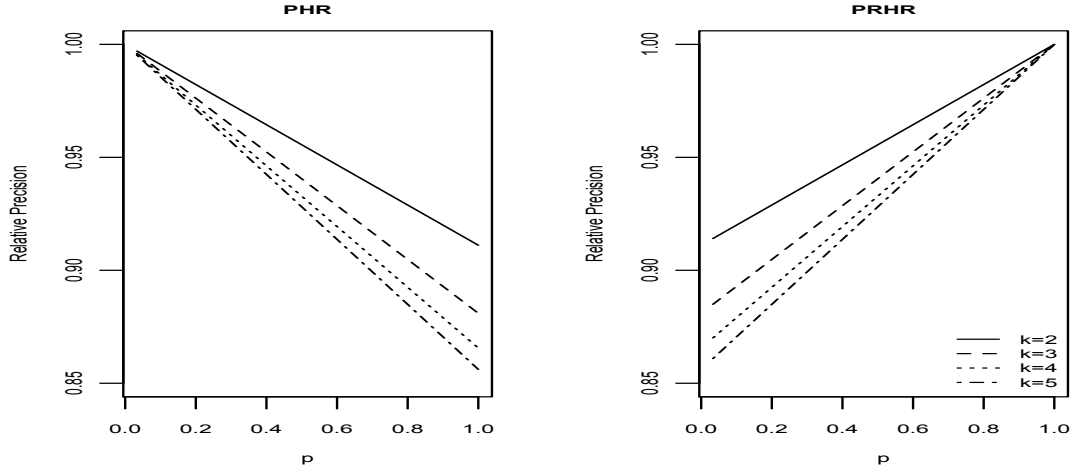


Figure 4: The relative precision of the MM estimators of $\gamma(\theta)$ based on the RNS complete-data over their SRS counterparts in the PHR (left panel) and PRHR (right panel) models when $k_i = 2, 3, 4$, and 5, and the proportion of maximums is $p = 0.1(0.1)1$. The relative precision less than one shows RNS performs better than SRS.

Looking at the top right panel, which presents the performance of the MM estimator of λ in the RNS incomplete-data Type II, shows that RNS designs with design parameters ρ_1 , ρ_2 , and ρ_3 do not perform better than SRS for all the examined parameters λ . For ρ_4 , the performance of the RNS incomplete-data Type II improves over SRS. The next four panels in the middle and down in Figure 5 present the performance of the MM estimators of λ in the RNS incomplete-data Type II for $\zeta \in \{0, 0.25, 0.75, 1\}$, where $\zeta = 1$ has the best impact on the performance of this type of the RNS design especially when the parameter value λ increases. For small λ the proportion of SRS, samples from the sets of size $K = 1$, should be small.

In Figure 6, the performance of the RNS-based MM estimators of η in the introduced beta distribution are compared with their corresponding estimators in the SRS design. To evaluate the relative precision in the RNS incomplete-data Type I, the RNS-based MM estimators of $\eta = 1, 2, 3, 4$ for $\zeta \in [0, 1]$ are examined. The top left panel shows that, no matter what the value of η is, $\zeta = 0$ provides the best performance of the RNS incomplete-data Type I compared with the SRS scheme. Considering the top right panel, which shows the relative precision of the RNS-based MM estimators in the incomplete-data Type II, it is seen that for all assumed distributions on K with larger and fixed set sizes, i.e., ρ_4 and regardless of the parameter value η , the RNS outperforms SRS. The next panels in Figure 6 confirm that $\zeta = 0$ provides the most efficient RNS-based MM estimators of η in the incomplete-data Type III. As ζ increases, for all values of η the performance of this type of RNS gets worse.

5.2 A Case Study

Fish products have been shown to contain varying amounts of heavy metals, particularly mercury from water pollution. Mercury is dangerous to both natural ecosystems and humans because it is a metal

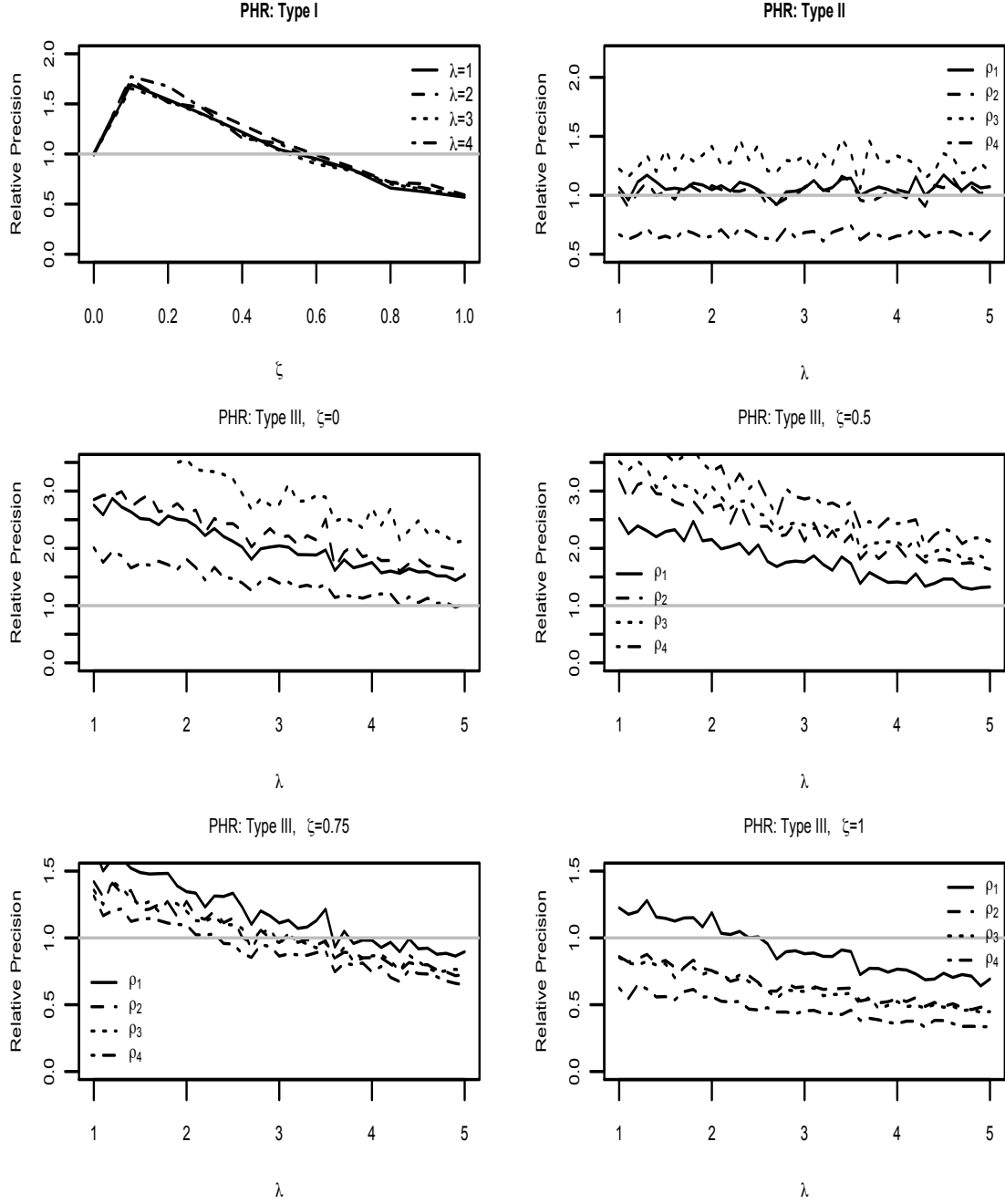


Figure 5: The relative precision of the RNS incomplete-data Type I (top left panel) for $\zeta \in [0, 1]$ and $\lambda \in \{1, 2, 3, 4\}$, Type II (top right panel) for four distributions on K and $\eta = 1(0.1)5$, and Type III (middle and lower panels) for $\zeta \in \{0, 0.25, 0.75, 1\}$ and $\eta = 1(0.1)5$ in a beta distribution with parameter $1/\eta$, the shape parameter $\beta = 1$, and $m = 10$. Values of the relative precision less than one shows RNS performs better than SRS.

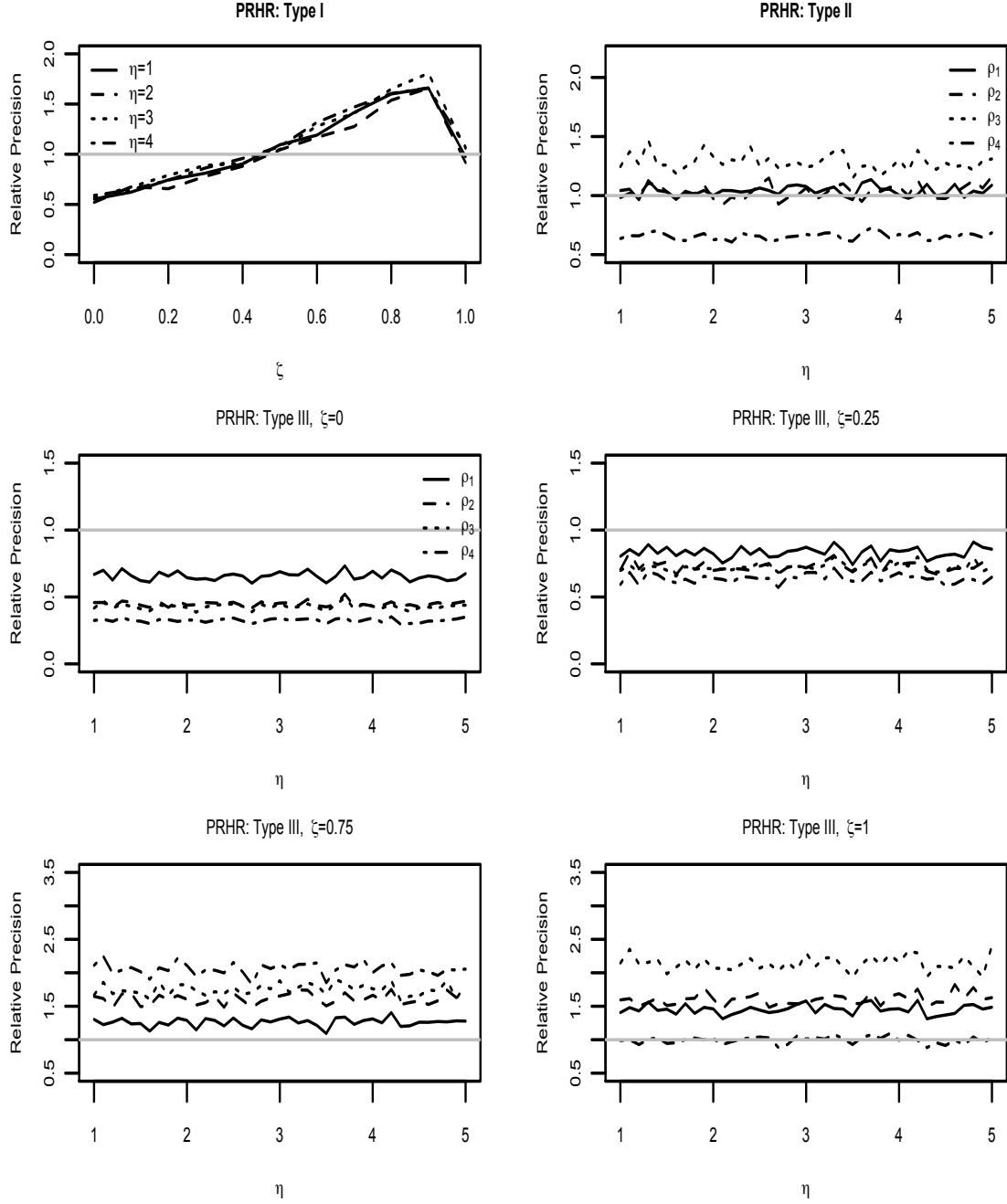


Figure 6: The relative precision of the RNS incomplete-data Type I (top left panel) for $\zeta \in [0, 1]$ and $\eta \in \{1, 2, 3, 4\}$, Type II (top right panel) for four distributions on K and $\eta = 1(0.1)5$, and Type III (middle and lower panels) for $\zeta \in \{0, 0.25, 0.75, 1\}$ and $\eta = 1(0.1)5$ in a beta distribution with parameter $1/\eta$, the shape parameter $\beta = 1$, and $m = 10$. Values of the relative precision less than one shows RNS performs better than SRS.

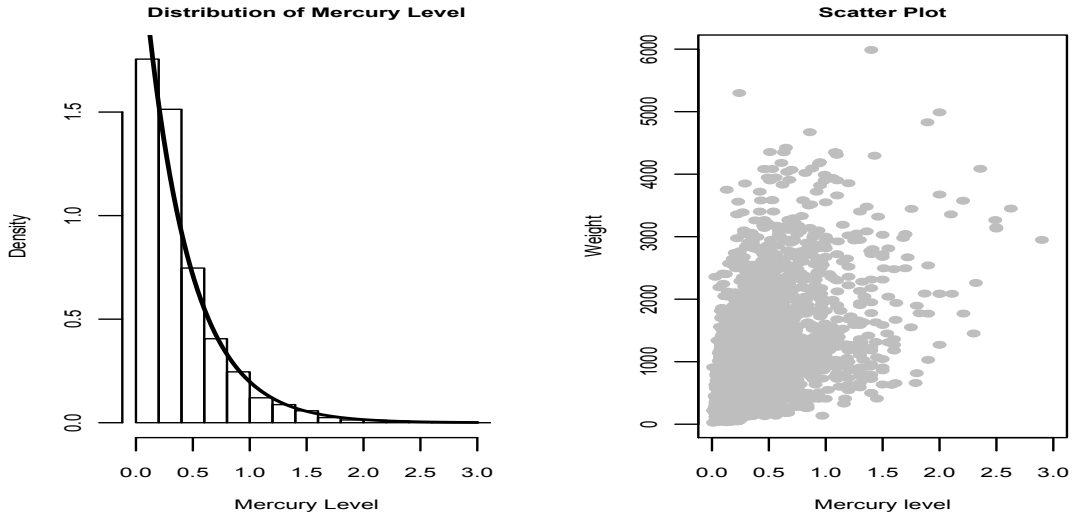


Figure 7: The population shape of the fish mercury level (left panel) and scatterplot between the fish mercury level and fish weight.

known to be highly toxic, especially due to its ability to damage the central nervous system. Children, as well as women who are pregnant or planning to become pregnant, are the most vulnerable to mercury’s harmful effects. Many studies have been performed to determine the mercury concentration in the fish species and evaluate the performance of the proposed remedial actions (e.g. Bhavsar et al. (2010), McGoldrick et al. (2010) and Nourmohammadi et al. (2015a)).

The mercury grows in concentration within the bodies of fish. It is well-known that measuring the mercury level in fish needs an expensive and complex laboratory procedure. However, a small group of fish can be ranked based on their mercury contamination levels before taking the actual measurements on them. This can be done using either the fish weight or length which have positive correlations with mercury contamination levels in fish. In this section, the performance of the RNS-based modified maximum likelihood estimator of the distribution parameter is compared with its corresponding SRS counterpart. The dataset contains mercury levels, weights, and lengths of 3033 *sandra vitreus* (Walleye) caught in Minnesota. The original database contains 102,850 fish tissue mercury records compiled by the United States Geological Survey from various sources. Selected records were those that: had a non-zero mercury concentration; had a valid species identifier; had non-zero length and weight measurements; and had a sample type of “filet skin on”.

In this study both perfect and imperfect ranking settings are considered. In the perfect ranking setting the study variable, i.e. the fish mercury level, is used for ranking the sample units in the sets and for the imperfect ranking setting, ranking is performed using the fish weights. Kendall’s τ between the mercury level and weight values is about 0.4, which is relatively small. Figure 7 (left panel) shows the population shape for the mercury levels of 3033 fish records and it looks the fish mercury level follows an exponential distribution, a member of PHR model, with parameter $\lambda = 0.3902$. The choice of exponential distribution for fish mercury level is justified by the Kolmogorov-Smirnov test. Figure 7 (right panel) shows the scatterplot between the mercury level and weight. The RNS complete-data

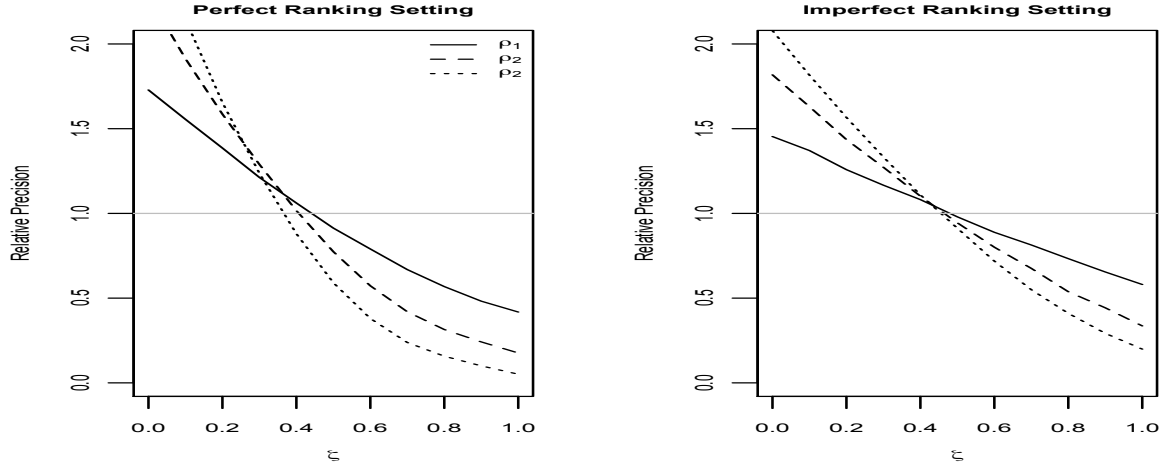


Figure 8: The Relative precision of the ML estimators of λ based on RNS incomplete-data Type III over their SRS counterparts in the perfect ranking setting (left panel) and imperfect ranking setting (right panel) when the parameters ρ and ζ are assumed to be unknown for $\zeta = 0(0.1)1$ and ρ_1, ρ_2, ρ_3 .

and three RNS incomplete-data scenarios were considered in section 5.1. In the examined three types RNS incomplete-data, the parameters ρ and/or ζ are assumed to be known. In this section we examine the performance of RNS incomplete-data Type III in which the parameters ρ and ζ are unknown. The performance of the design in estimating the population parameter is evaluated using the relative precision, i.e. the mean square of RNS-based MLE of λ over the mean square of its SRS counterpart. The relative precision is calculated for the MLE's of λ for $\zeta = 0(0.1)1$ and three distributions of the set size K as follow

$$\rho_1 = (0.4, 0.3, 0.2, 0.1), \quad \rho_2 = (0.1, 0.2, 0.3, 0.4), \quad \rho_3 = (0, 0, 0, 1).$$

Figure 8 provides the relative precision of the ML estimators of parameter λ based on the RNS incomplete-data Type III. It shows that $\zeta = 1$, regardless of the parameter value ρ , is the optimum value of ζ which improves the RNS-based ML estimator in the incomplete-data Type II scenario compared with their SRS counterparts. The relative precision of RNS-based estimators presented in Figure 8 is obtained in an EM algorithm when the initial values of ρ and ζ are $\rho_0 = (0.25, 0.25, 0.25, 0.25)$ and $\zeta_0 = 1$. Considering the sensitivity of the EM algorithm to the initial values of the unknown parameters ρ and ζ , we also examined $\rho_0 = (1, 0, 0, 0)$, $\rho_0 = (0, 0, 0, 1)$, $\zeta_0 = 0$ and $\zeta_0 = 0.5$. Except $\rho_0 = (1, 0, 0, 0)$ and $\zeta_0 = 0$, for the other examined initial values of ρ and ζ , RNS outperforms SRS for larger values of true ζ 's, i.e. $\zeta \in (0.5, 1]$, and it shows the best performance of RNS over SRS at $\zeta = 1$.

6 Concluding remarks

Randomized nomination sampling (RNS) was introduced by Jafari Jozani and Johnson (2012) and it has been shown to perform better than simple random sampling (SRS) in constructing nonparametric

confidence and tolerance intervals. RNS has potentials for a wide range of applications in medical, environmental and ecological studies. In this paper, we described the RNS-based ML and MM estimators of the population parameters when the underlying study variable follows PHR or PRHR model. Various conditions on the type of information, ranking error settings and the design parameters including distribution of the set size (ρ) and probability of taking the maximum observation of the set (ζ) have been investigated. We introduced four types of RNS data, corresponding to situations in which all the observations, set sizes and observations ranks in the sets are known, only observations and the set sizes are known, only the observations and their ranks in the sets are known, or finally only the observations are known. Considering all the situations, we also provided the PDF and CDF of an RNS observation. We showed that there is always a range of ζ on each RNS is superior to SRS in terms of the relative precision. The RNS design has this advantage regardless of the ranking setting. The relative precision of the estimators obtained in the RNS design becomes better when more weight is given to the larger set size and $\zeta = 1$ (in PHR model) or $\zeta = 0$ (in PRHR model).

Acknowledgements

The authors gratefully acknowledge the partial support of the NSERC Canada.

References

- Al-Odat, M. and M. F. Al-Saleh (2001). A variation of ranked set sampling. *Journal of Applied Statistical Science* 10(2), 137–146. 2
- Bhavsar, S. P., S. B. Gewurtz, D. J. McGoldrick, M. J. Keir, and S. M. Backus (2010). Changes in mercury levels in great lakes fish between 1970s and 2007. *Environmental science & technology* 44(9), 3273–3279. 22
- Boyles, R. A. and F. J. Samaniego (1986). Estimating a distribution function based on nomination sampling. *Journal of the American Statistical Association* 81(396), 1039–1045. 2, 3
- Gemayel, N. M., E. A. Stasny, and D. A. Wolfe (2010). Optimal ranked set sampling estimation based on medians from multiple set sizes. *Journal of Nonparametric Statistics* 22(4), 517–527. 3
- Ghosh, K. and R. C. Tiwari (2009). A unified approach to variations of ranked set sampling with applications. *Journal of Nonparametric Statistics* 21(4), 471–485. 2
- Gupta, R. C., P. L. Gupta, and R. D. Gupta (1998). Modeling failure time data by lehman alternatives. *Communications in Statistics-Theory and Methods* 27(4), 887–904. 4
- Gupta, R. C. and R. D. Gupta (2007). Proportional reversed hazard rate model and its applications. *Journal of Statistical Planning and Inference* 137(11), 3525–3536. 4

- Helsen, K. and D. C. Schmittlein (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science* 12(4), 395–414. 4
- Jafari Jozani, M. and B. C. Johnson (2012). Randomized nomination sampling for finite populations. *Journal of Statistical Planning and Inference* 142(7), 2103–2115. 2, 23
- Jafari Jozani, M. and S. J. Mirkamali (2010). Improved attribute acceptance sampling plans based on maxima nomination sampling. *Journal of Statistical Planning and Inference* 140(9), 2448–2460. 2
- Jafari Jozani, M. and S. J. Mirkamali (2011). Control charts for attributes with maxima nominated samples. *Journal of Statistical Planning and Inference* 141(7), 2386–2398. 2
- Kvam, P. H. and F. J. Samaniego (1993). On estimating distribution functions using nomination samples. *Journal of the American Statistical Association* 88(424), 1317–1322. 2
- Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, Volume 362. John Wiley & Sons. 4
- McGoldrick, D. J., M. G. Clark, M. J. Keir, S. M. Backus, and M. M. Malecki (2010). Canada’s national aquatic biological specimen bank and database. *Journal of Great Lakes Research* 36(2), 393–398. 22
- Mehrotra, K. and P. Nanda (1974). Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika* 61(3), 601–606. 5
- Navarro, J., F. J. Samaniego, N. Balakrishnan, and D. Bhattacharya (2008). On the application and extension of system signatures in engineering reliability. *Naval Research Logistics (NRL)* 55(4), 313–327. 4
- Nourmohammadi, M., M. Jafari Jozani, and B. C. Johnson (2014). Confidence intervals for quantiles in finite populations with randomized nomination sampling. *Computational Statistics & Data Analysis* 73, 112–128. 2
- Nourmohammadi, M., M. Jafari Jozani, and B. C. Johnson (2015a). Distribution-free tolerance intervals with nomination samples: Applications to mercury contamination in fish. *Statistical Methodology* 26, 16–33. 2, 22
- Nourmohammadi, M., M. Jafari Jozani, and B. C. Johnson (2015b). Nonparametric confidence intervals for quantiles with randomized nomination sampling. *Sankhya A* 77(2), 408–432. 2
- Samawi, H. M., M. S. Ahmed, and W. Abu-Dayyeh (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal* 38(5), 577–586. 2
- Tiwari, R. C. (1988). Nonparametric bayes estimation of a distribution under nomination sampling. *Reliability, IEEE Transactions on* 37(5), 558–561. 2

- Tiwari, R. C. and M. T. Wells (1989). Quantile estimation based on nomination sampling. *Reliability, IEEE Transactions on* 38(5), 612–614. 2
- Wells, M. T., R. C. Tiwari, et al. (1990). Estimating a distribution function based on minima-nomination sampling. In *Topics in statistical dependence*, pp. 471–479. Institute of Mathematical Statistics. 2
- Willemain, T. R. (1980). Estimating the population median by nomination sampling. *Journal of the American Statistical Association* 75(372), 908–911. 2